1959

Statistical sampling as an auditing tool

Oscar S. Gellein

Statistical Sampling as an Auditing Tool

By Oscar S. Gellein
Partner, Newark Office

Presented at the Annual Meeting of the American Institute of Certified Public Accountants, San Francisco — October 1959

Statistical sampling requires selection of the items making up the sample in such a way that the *laws of probability* determine the manner of occurrence of the items in the sample. When the selection is carried out in this way, measurements can be made with respect to the sample and mathematical inferences drawn from them. The measurements are called precision and reliability (or level of confidence). The sense of these measurements is not alien to our intuition.

We know intuitively that there is one chance in two that a tossed coin will show a head. We also know intuitively that if a coin is tossed a large number of times it is likely that a head will appear approximately 50 per cent of the time. We are prudent enough, however, to avoid saying with certainty that a head will appear, say, 5,000 times out of 10,000 tosses. I am sure, however, that our intuition tells us that the greater the number of tosses the greater the reliability, or the confidence, that a head will appear one-half the time.

Now statistical sampling adds another dimension which also falls within the purview of our intuition. It says that with certain precision for a given degree of reliability a statement can be made about the number of heads that will appear if a coin is tossed, say, 100 times. In other words, if the statement is to be made with 95 per cent reliability it must refer to a range around 50—in this case about 40 to 60—the precision would be ± 10, the reliability 95 per cent. Another way to say it is that if, day after day, 100 persons each made 100 tosses of a coin, on the average, 95 of them would show a number of heads falling in the range of 40 to 60. These results are calculated, not estimated. They can be stated with mathematical certainty.

We could look at this situation another way, too. Suppose that we had no previous knowledge about a coin, that is, about the number of times that a head might be expected to appear in any given number of tosses. Could we in these circumstances, after tossing it a number of times, make any statement about this feature of a coin? I think

again our intuition says that we could. Only this time our statement would have to say that with a given degree of reliabiliy there is a measurable range of precision containing the actual value. If 48 heads appeared, for example, in 100 tosses, we would say that we are 95 per cent confident that a coin is constructed in such a way that the range of 38% to 58% contains the per cent of heads that will appear in an extremely large number of tosses. In a sense, this parallels the usual sampling situation where one seeks to learn about some feature of the whole from which the sample is drawn.

SAMPLE SIZE AND PRECISION AND RELIABILITY

In a capsule, then, statistical sampling approaches the matter by saying that if items are selected for the sample in such a way that all items have equal chances of being included, the laws of probability are at work. Further, when the laws of probability are working certain mathematical inferences can be drawn about the whole from which the sample was drawn. The inferences cannot be stated, obviously, with complete assurance of accuracy, but they can be stated with measurable reliability for a measurable range of precision. This is the nub of it—measurable reliability and precision. If greater reliability is desired for given precision, or tighter precision for given reliability, a larger sample may be required—this, too, is a conclusion that we would reach intuitively.

Two observations at this point may be worth while. It is demonstrable that the absolute size of the sample is more meaningful than the relative size in considering precision and reliability. This may not, at least at first glance, square with our intuition. But ordinarily where one group of items to be sampled is twice as large, for example, as another group it is not necessary to take a sample twice as large for the same precision and reliability. In many cases it may be substantially less than twice as large.

The other observation is that there is an indefinite number of combinations of precision and reliability in any situation for any given sample size. For example, in tossing a coin 100 times we could say with 95 per cent reliability that a head would appear anywhere from about 40 to 60 times but we could say with only 68 per cent reliability that it would appear anywhere from about 45 to 55 times. The finer the bead is expected to be drawn, the lower the confidence.

RANDOMNESS

All of this is upset, however, if there is some feature of the coin or if it is tossed in some way or if it falls in some way that causes the appearance of either a head or a tail to be favored. Since on each toss a head is as likely to appear as a tail, and conversely, the laws of probability are operating. In more general terms it can be said that the toss of a coin furnishes a random selection of one of its faces. A random selection is not the same thing as a haphazard selection. Haphazard selections may be affected by human predilections. Randomness requires removal of all factors that might favor the selection of any item. Randomness is ordinarily achieved by the use of random number tables, a number of which are readily available.

My simplified observations on the essential ingredients of statistical sampling could, of course, be misleading unless they are viewed as comments intended only to deal in a general, nontechnical way with some of the principal characteristics of this type of sampling. They are not intended to imply, either, that statistical sampling in all of its forms is simple. The breadth and depth of it, as a matter of fact, is substantial.

Statisticians have developed various methods, some of them quite sophisticated, for dealing with many different types of situations in which partial data may be drawn and the results evaluated for particular purposes. It has been my intention to point out, however, that the identifying characteristic, the unique feature, of all of them, concerns randomness in selection which, in turn, permits measurement of precision and reliability.

SURVEY SAMPLING, ACCEPTANCE SAMPLING, DISCOVERY SAMPLING

Three different forms of statistical sampling have been recommended for use in connection with various accounting and auditing matters. When I say recommended for use I refer both to the actual applications that have been written up in the literature and to items in the literature dealing with abstract considerations.

First, there is sampling intended to furnish a calculation of some feature of the whole from which the sample is taken. It may be termed *survey sampling.* The feature to be calculated might be the

per cent of errors of a particular type in the whole, or the dollar amount of the whole, or the average amount of the items making up the whole. I shall refer later to some applications of survey sampling in accounting.

Second, there is acceptance sampling. The aim here is to develop a sampling method that has built into it the basis for an automatic decision depending upon the results of the sample. For example, it might be decided that some other auditing procedure will be applied if the rate of errors in a sample exceeds, say, 2 per cent and, concurrently, that nothing further would be done if the rate is 2 per cent or less. The elements of precision and reliability are present in this situation as in other forms of statistical sampling. The risk (the complement of reliability) that the decision was based upon misleading sample information is measurable here, too.

The acceptance sampling approach is akin to quality control, the statistical method which, as you know, has been used extensively in connection with product inspection in manufacturing. Some auditors feel that this approach has limited application to independent auditing mainly because the auditing process just does not work in an accept-reject way. Others feel that it may be useful to the auditor.

The third form of statistical sampling to attract the interest of those considering possible applications to auditing is discovery sampling. Discovery sampling is designed to give desired reliability that the sample will include at least one item of a particular kind if the number of items like it in the whole being examined is as large as some designated figure. For example, this approach might be used if one wished to have a particular degree of confidence that he would find at least one case of a misfooting if there were as many as, say, fifteen misfootings.

A considerable number of tables have been prepared for acceptance sampling and for certain applications of survey sampling. They are intended to permit determination of the sample size for varying sizes and conditions of the whole from which the sample is taken and for varying degrees of reliability. They are intended to facilitate the sampling process. Obviously, there are some applications for which it is necessary to use the formulas underlying the tables.

The differences between these types of sampling relate only to their form. There is, for example, a feature of estimating in each of them. They have the same substance, or basis—operative laws of probability because of randomness.

A quick look at an application to accounting may be helpful. In 1958 when the AICPA committee on statistical sampling prepared a selected bibliography of articles and books relating to applications of statistical sampling to accounting and auditing, about a dozen pages were required for it. A number of these publications describe or refer to applications that have been made either experimentally or as adopted practice.

In one such case the work-in-process inventory at one of the divisions of a company was determined by statistical sampling. This was an application of survey sampling where the dollar amount of the physical inventory was being determined. The company reports that it carried out extensive studies and experimentation over several years before putting the method into effect. The characteristics of the inventory, the variation of the dollar amounts of the lots making up the inventory, the problems in making a random selection of lots in a factory, the matter of training personnel, writing instructions, and all similar matters were thoroughly studied. Pilot studies were made in selected departments to ascertain problems of implementation. In addition, a year ahead of starting the method complete simulation of it based upon IBM cards representing the inventory, on a card-for-lot correspondence, was undertaken.

Consideration of the distribution of the 40,000 lots making up the inventory led to the conclusion that the greatest sampling efficiency would be achieved with a 100 per cent count of the high-valued lots. An essential preliminary consideration, of course, was the combination of reliability and precision desired in the circumstances. This decision and the one relating to the complete count of the high-valued lots, in turn, led to the determination of the sample size relating to the low-valued lots. The result was a count of 4,200 of the 40,000 lots —all of the high-valued lots and a random sampling of the low-valued lots. The selection of the low-valued lots was made by use of a random number table. The value of the inventory as a whole was then computed from the results of the sample and the count of the high-valued items.

Among the advantages claimed for this method by the company official reporting on it were the following:

- Greater accuracy, because tighter controls and greater skills could be applied to fewer counts.

- It produced an inventory figure to which could be ascribed known limits of precision and known reliability.
- A considerably shorter period of time was required for inventory-taking permitting better dovetailing of vacation schedules of all personnel with the plant shutdown.
- There was lower cost of inventory-taking (this was not considered by the company to be the most important advantage).

As a further test of the quality of the work done in counting and identifying the lots included in the survey sample a random sample of these lots was drawn on an acceptance sampling basis. This was used to decide whether recounts were necessary.

OTHER APPLICATIONS

Among the other applications of survey sampling to accounting information referred to in the literature are those concerned with: determining the price index to be applied in the dollar valuation of a Lifo inventory; determining the adjustment of the perpetual inventory total necessary to conform it with the physical inventory total (this was done by sampling for an average dollar adjustment per item), and determining interline settlements between transportation companies.

In recent years both the Auditor General's office of the U. S. Air Force and the U. S. Army Audit Agency have given considerable attention to statistical sampling. The literature contains descriptions of some of their work and some of their conclusions. In some of their applications they have used survey sampling in measuring rates of errors in preparing various records. It might be said that their purpose is more nearly that of the internal auditor of a company than of the independent auditor examining financial statements. In one application concerned with supply activities at Air Force installations random samples were drawn from vouchers showing receipts, issues, adjustments, and other related operations. They first listed the kinds of errors that might occur in a voucher, estimated sample sizes based on all information available to them, selected vouchers randomly, applied their usual procedures in checking out the vouchers for errors, and, based on the statistical estimates of error rates, took whatever action seemed necessary to tighten controls.

So far I have said little about applications of statistical sampling to the work of the independent auditor in examining financial statements as the basis of his opinion as to their fairness. I have done so because it is desirable to recognize that the purposes of original accounting, of internal auditing, and of independent auditing, although related, are significantly different. Statistical sampling, its form and method, is geared to the purposes to be achieved in studying the items to be tested. Different forms of sampling are applied to achieve varying purposes. For example, it seems not unlikely to me that applications of acceptance sampling might be more frequent in internal auditing than in independent auditing, because of the greater concern in internal auditing with the efficiency of clerical activities as a continuing stream of activities. This is not to imply, of course, that the independent auditor is not concerned with clerical efficiency.

I should like to make still another delineation clear. I am not inclined to use the terms "sampling" and "testing" interchangeably. Testing to me implies, of course, looking at partial data. I think it implies, in addition, the judgment the auditor applies in selecting items, in deciding what audit work to do with the items selected, in doing the work, and in evaluating the results of it. This is not to say that he may not in some circumstances wish to make use of statistical sampling in making his selection of items for the test. In some circumstances, too, he may wish to make use of the evaluation phase of statistical sampling to arrive at measures of precision and reliability. In other words, I think statistical sampling is a tool that should be available to the independent auditor along with all of the other techniques available to him in gaining the satisfaction that he requires for the expression of an opinion. He may wish to use it in circumstances where he thinks that it furnishes either an improved basis of selecting items or some guides to the question of when he does the right amount of work.

AUDITORS' CONCERNS

As auditors we have a natural concern about a process that appears to place the audit in a mathematical mold capable of being used to produce automatic readings about the fairness of the financial statements we are examining. We fear that a slide rule will supplant our judgment, that the areas in which the auditor's judgment may be exercised will be restricted. I think these concerns are unfounded.

In the first place there is nothing in statistics saying that judgmental selection of partial data should always be replaced by statistical sampling. Efficient use of statistical sampling requires a clear-cut operative statement of the purpose of the test. We know that in auditing to a large extent we do a lot of things to express a composite opinion on financial statements. We know that as to any one thing that we do we may gain satisfaction about several items in the financial statements and that as to any one item in the financial statements we gain satisfaction by doing several things. We know that a given test may have several purposes. For example, in comparing duplicate copies of deposit slips with the cash records we may be looking for signs of lapping or for signs of whether receipts are being deposited promptly. These purposes in turn fuse into and with others in arriving at a composite opinion as to internal control. Further, we do other things to gain satisfaction about similar matters. For example, confirmation by direct correspondence with customers has some of the same purposes as comparison of deposit slips with cash records. All I am saying is that the audit process is such that as to any one step the auditor may not be able to isolate it and say that it has a particular purpose. In another engagement, too, the same step may have another mixture of purposes.

AUDITORS' JUDGMENT

I must say still another word about this matter of statistical sampling's supplanting the auditor's judgment. It need not do so, if used wisely—instead, it should supplement it. For example, there is nothing in the statistical application that deters the auditor from looking at all items of significant dollar-amount. Instead statistical sampling would do the same thing by stratification to achieve the purpose the auditor has in mind. There is nothing in statistics deterring the auditor from, for example, confining his tests to a particular day if there is an audit purpose to be achieved in doing so. There is nothing in statistics that will force the auditor to accept a purpose of a test incompatible with the auditing process. In addition, it is for the auditor to decide what precision and reliability he is seeking in a test.

Further, there is nothing in statistics requiring the auditor to change the nature of the work that he will do with the items selected or to alter the actions that he will take in connection with his tests. There is nothing in statistics, either, that requires the use of one form

of statistical sampling in preference to another. The audit purpose, instead, is controlling in this regard. The essential decisions are made by the auditor: defining the purpose of the test, selecting the audit procedures to be applied, and fixing the reliance sought.

Pursuit of consideration of the place that statistical sampling might hold in auditing probably will have salutary effects in other ways. It will require us to put in fairly sharp focus the purpose of many of our auditing procedures. It may reveal to us, for example, that in some circumstances testing footings for particular months may not be achieving the purpose as efficiently as selecting on a random basis throughout the year the pages for which footings are to be checked. It may demonstrate some situations, for example, that in selecting vouchers for examination or paid checks for inspection the audit purpose may better be served by random selection throughout the year than by selection of particular months for which the items will be examined. Such questions will have to be resolved, in part, on the basis of deciding whether we are seeking to learn about how internal control has been working throughout the year or how it has been working during the particular months chosen.

ADVANTAGES TO THE AUDITOR

These problems and all of the others concerned with gearing sampling to audit purposes should not, I think, deter us from pursuing our consideration of the places in independent auditing where statistical sampling might facilitate achieving audit purposes more efficiently. In those areas of the audit where statistical sampling is applicable it may have these advantages to the auditor:

- His intuitive conclusions would be bolstered in some instances.
- In some instances he would be deterred from drawing unwarranted inferences.
- He would have some guides as to when he has done the right amount of work, not too much and not too little.

RELATIVELY SLOW DEVELOPMENT

Developments in applying statistical sampling to auditing have been slower than in applying them to internal accounting. The reasons I think relate to the matters to which I have referred: mixture of purposes, interrelationships of audit procedures, and difficulties in

defining errors in such a way as to make meaningful measures of precision and reliability concerning their presence.

One feature of the audit process requires considerable study, as I see it, before any general conclusions can be reached about extensive application of statistical sampling to independent auditing. It relates to the combined reliance that we place on internal control to furnish reasonable records and on our tests of the records themselves. There is another leg, too, of this triangle—some of the testing is done to gain satisfaction about internal control. In other words part of our reliance—and I think you can substitute the statistical term *reliability*—is subjectively gained. We size up the people working with the records, we see how they are organized and supervised, and we make inquiries of them and of others. In many instances a significant part of our reliance stems from subjective evaluations. The reliability we seek in our tests in some way must, accordingly, be combined with the subjective evaluation.

On the other hand, these problems are not of such nature as to preclude completely the application of statistical sampling from independent auditing, and I intend no such implication. As a matter of fact, some such applications are described in the literature. One concerns confirmation of accounts receivable in a department store to estimate the proportion of errors in the accounts as a whole. Another concerns a testing of the aging of accounts receivable. It is not possible for me to appraise the extent to which the applications are being made in independent auditing. I think it a fair assessment to say that some have been made and have been described in the literature, others undoubtedly have been made on a recurring basis, and that quite a good deal of investigation and experimentation is under way.

TAKING UP A STUDY OF STATISTICAL SAMPLING

How does an auditor go about getting his study of statistical sampling under way? In the first place there is a fairly significant body of literature dealing with both the theoretical considerations and practical applications with which he can become acquainted. The second step, it seems to me, requires teamwork application of a knowledge of auditing and a knowledge of statistics to a study of audit engagements for the purpose of identifying the tests where a relatively large number of items is to be tested, where the auditing purpose is clearly defined, where the auditor's satisfaction would be bolstered

or supplemented by knowledge of the precision and reliability of his test, and where the statistical knowledge leads to the conclusion that the auditing purpose can better be served by statistical sampling. I put it this way because certified public accountants are going to find different ways of developing this teamwork. In some cases the CPA himself may, by reason of previous training, be able to acquire sufficient knowledge of statistics to be able to do it himself. In most cases, however, he probably will have to obtain the services of a statistician either by employing or by engaging him on a consulting basis, possibly in a pool arrangement with other accountants. In my view, the initial consideration cannot be given satisfactorily either by the auditor with little or no knowledge of statistics or by the statistician with little or no knowledge of auditing. It does not follow that every CPA must be an expert statistician. After the initial stage of study has been completed, a significantly lesser knowledge of statistics may be sufficient to permit applications to be carried out. If this does not prove true, the scope and extent of applications of statistical sampling to auditing likely will be rather narrowly limited for quite some time.

In this teamwork stage there must, it seems to me, be experimental applications to learn more about mechanical problems of application, to get a feeling for meaningful precision ranges and reliability (or confidence limits), and to discover whether statistical sampling might have made more efficient and meaningful the tests that the auditor has been applying. Developments in the teamwork stage will undoubtedly lead to the charting of the next stage.

There is no ready-made do-it-yourself kit for applying statistics to auditing. There is no short-cutting to be done. The whole matter is so central to auditing, since it deals with auditing purposes of procedures and tests, that every CPA must in varying degrees go through this process of teamwork study.